AD-A161 963 TESTS FOR INITIALIZATION BIAS IN COMPUTER SIMULATION EXPERIMENTS(U) CORNELL UNIV ITHACA NY SCHOOL OF OPERATIONS RESEARCH AND INDU D GOLDSMAN ET AL. DEC 84 TR-J-84-16 N00014-81-K-0037

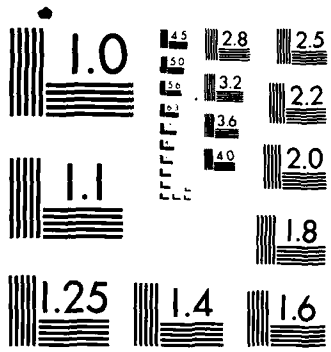
1/1

UNCLASSIFIED F/G 9/2 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

School of Industrial and Systems Engineering

Georgia Institute of Technology

Atlanta, GA 30332

Technical Report No. J-84-16

December, 1984

Tests for Initialization Bias in Computer Simulation Experiments

by

David Goldsman

and

Lee Schruben

This research was partially supported by the Office of Naval Research under contract N00014-81-k-0037. Lee Schruben is currently in the School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853.

Abstract

Although many of the rules for detecting and dealing with *initialization bias* in computer simulation experiments are easy to understand and implement, they are nonetheless heuristic. ~~The current~~ This paper uses the theory of standardized time series to construct tests which (under certain conditions) detect "significant" initialization bias in a process. Previous tests for initialization bias can be viewed as *special cases* of the general family of tests to be presented here.

1. INTRODUCTION

When a system is simulated, it is sometimes difficult to select appropriate initial conditions to drive the simulation. We often wish to initialize the process with "typical" system values; however, if there is little *a priori* knowledge of the process available, we might initialize the system in a state which has a very low probability of occurring. This *initialization bias* (or *initial transient*) can then result in incorrect conclusions on the part of the experimenter.

A common way of dealing with the initial transient is to take a very large (possibly wasteful) number of observations - large enough so that the initialization effects are overwhelmed. Perhaps a better method of counteracting the bias problem is simply to delete (truncate) a portion of the output from the beginning of the simulation run. The experimenter would then hope that the offending biased observations had been eliminated. Unfortunately, if the output is truncated too early, then significant initialization bias might still be present. If it is truncated too late, then "good" observations are lost. [cf. Snell and Schruben (1979).]

Although many of the rules for detecting and dealing with initialization bias are easy to understand and implement, they are nonetheless heuristic [see, e.g., the surveys by Wilson and Pritsker (1978) and Schruben and Goldsman (1984)]. The current paper uses the theory of *standardized time series* to construct tests which (under certain conditions) detect "significant" initialization bias in a process. Previous tests from Schruben (1982) and Schruben, Singh, and Tierney (1983) can be viewed as *special cases* of the general family of tests to be presented here.

This paper is organized as follows. Background material is

provided in Section 2. Our new tests are introduced in Section 3. Section 4 is concerned with power calculations for the new tests.

2. PRELIMINARIES

2.1 Some Standardized Time Series Results

Consider the stochastic process $X_1,\ldots,X_m$. For $j=1,\ldots,m$, let $\bar{X}_j \equiv \frac{1}{j}\sum_{i=1}^{j} X_i$ and $S_j \equiv \bar{X}_m - \bar{X}_j$; also, define $\sigma^2 \equiv \lim_{m\to\infty} m\mathrm{Var}(\bar{X}_m)$. The *standardized time series* is $T_m(t) \equiv \dfrac{\lfloor mt \rfloor S_{\lfloor mt \rfloor}}{\sigma\sqrt{m}}$, $0 \le t \le 1$, where $\lfloor\cdot\rfloor$ is the greatest integer function.

Suppose that $X_1,\ldots,X_m$ is a sequence of stationary, finite variance random variables. In Schruben (1983), it is shown that under rather mild assumptions, $T_m(t) \xrightarrow{\mathcal{D}} B_t$ as $m \to \infty$, where B_t is a standard Brownian bridge process. Further, $T_m(t)$ is asymptotically independent of $m\bar{X}_m$.

Let us divide the stationary, finite variance series $X_1,\ldots,X_n$ into b adjacent *batches*, each consisting of m X_j's $(n=bm)$; the random variables $X_{(i-1)m+1}, X_{(i-1)m+2}, \ldots, X_{im}$ comprise batch i, $i=1,\ldots,b$. If m is large enough, we can treat the batches as if they were (approximately) independent. Each individual batch can then be standardized; this yields b standardized time series which are approximately independent Brownian bridges.

For $i=1,\ldots,b$ and $j=1,\ldots,m$, define the random variables:

$$\bar{X}_{i,j} \equiv \frac{1}{j}\sum_{p=1}^{j} X_{(i-i)m+p} \quad \text{(cumulative averages).}$$

(Observe that $\bar{X}_{i,m}$ is the random variable corresponding to the i-th *batched mean*.)

$$\bar{\bar{X}}_n \equiv \frac{1}{b}\sum_{i=1}^{b} \bar{X}_{i,m} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{(grand mean)},$$

$$S_{i,j} \equiv \bar{X}_{i,m} - \bar{X}_{i,j},$$

$$\hat{K}_i \equiv \operatorname{argmax}_k \{kS_{i,k}\},$$

$$\hat{S}_i \equiv \hat{K}_i S_{i,\hat{K}_i}, \text{ and}$$

$$\hat{A}_i \equiv \sum_{j=1}^{m} jS_{i,j}.$$

<u>Theorem</u>: We have the following collection of estimators for σ^2:

(0) <u>Classical batched means estimator</u>:

$$V_{0,b} \equiv \frac{Q_{0,b}}{b-1}, \text{ where}$$

$$Q_{0,b} \equiv m\sum_{i=1}^{b}\left[\bar{X}_{i,m} - \frac{1}{b}\sum_{j=1}^{b} \bar{X}_{j,m}\right]^2 \xrightarrow{\mathcal{D}} \sigma^2\chi^2(b-1),\ b > 1.$$

(1) <u>Area estimator</u>:

$$V_{1,b} \equiv \frac{Q_{1,b}}{b}, \text{ where}$$

$$Q_{1,b} \equiv \frac{12}{m^3-m}\sum_{i=1}^{b} \hat{A}_i^2 \xrightarrow{\mathcal{D}} \sigma^2\chi^2(b),\ b \geq 1.$$

(2) <u>Combined classical-area estimator</u>:

$$V_{2,b} \equiv \frac{Q_{2,b}}{2b-1}, \text{ where}$$

$$Q_{2,b} \equiv Q_{0,b} + Q_{1,b} \xrightarrow{\mathcal{D}} \sigma^2\chi^2(2b-1),\ b > 1.$$

(3) Maximum estimator:

$$V_{3,b} \equiv \frac{Q_{3,b}}{3b}, \text{ where}$$

$$Q_{3,b} \equiv m \sum_{i=1}^{b} \frac{\hat{S}_i^2}{\hat{K}_i(m-\hat{K}_i)} \xrightarrow{D} \sigma^2\chi^2(3b),\ b \geq 1.$$

(4) Combined classical-maximum estimator:

$$V_{4,b} \equiv \frac{Q_{4,b}}{4b-1}, \text{ where}$$

$$Q_{4,b} \equiv Q_{0,b} + Q_{3,b} \xrightarrow{D} \sigma^2\chi^2(4b-1),\ b > 1.$$

Proof: See, e.g., Goldsman (1984).//

The above variance estimators will be used in this paper to test output from a stochastic process for initialization bias.

Remark: Other variance estimators arise from spectral methods [Heidelberger and Welch (1981)] and ARMA time series modelling methods [Fishman (1971,1973,1978) and Andrews and Schriber (1982)]; the current paper will not concentrate specifically on these estimators.

2.2 Previous Initialization Bias Tests

2.2.1 Motivation

Suppose that we model $X_1,\ldots,X_n$ as $X_i = \mu_i + Y_i$, $i=1,\ldots,n$, where $E[X_i] = \mu_i$ for all i, and the Y_i's are stationary. We say that *no initialization bias is present* if $\mu_i = \mu$, say, for all i. Otherwise, bias is present. Initialization bias might exist for higher order moments, but this case is not considered [see Schruben (1981)].

$\{\mu_i\}$ is the *transient mean process*. Figure 1(a) is representative of so-called *negative* initialization bias (as might be

encountered in a queue-length process when starting a system empty and idle). Figure 1(b) illustrates *positive* bias (e.g., inventory level after starting a system fully stocked). Figure 1(c) is a transient mean process which damps out. Processes (a), (b), and (c) each appear to be approaching "steady state"; this is indicative of initialization bias dying out as run-length increases. The process in Figure 1(d) has not yet approached steady state (and, in fact, may never).

It is unrealistic to expect that $\mu_i = \mu$ for all i. Hence, we will only be interested in detecting *significant* initialization bias.

The tests to be described in this paper can be motivated in the following ANOVA sense: We partition the process $X_1, \ldots, X_n$ into two contiguous, non-overlapping portions. For a particular realization of the process, a variance estimate based solely on the first portion of the output (or, alternatively, on the *entirety* of the output) is calculated. A variance estimate from the latter portion only of the output is also calculated. If the two variance estimates are deemed to be significantly different, then we reject H_0: $\mu_i = \mu$ for all i. [For the transient mean processes illustrated in Figures 1(a), (b), and (c), we would expect a variance estimate formed from the first portion of the output to be greater than an estimate from the latter portion.]

2.2.2 A Test Based on the Maximum Estimator

With the intent of applying the theory of standardized time series, Schruben (1982) assumes that the Y_i's satisfy the mild requirements alluded to in Section 2.1. The entire $\{X_i\}$ process is then standardized into one $\{T_n(t)\}$ process.

Under the "stationarity" hypothesis H_0: $\mu_i = \mu$ for all i, recall that the *maximum estimator* for variance (when the number of batches $b = 1$) is given by $V_{3,1} \approx (\sigma^2/3)\chi^2(3)$, where $\sigma^2 \equiv \lim_{n\to\infty} n\mathrm{Var}(\Sigma X_i/n) = \lim_{n\to\infty} n\mathrm{Var}(\Sigma Y_i/n)$ and the notation "$\approx$" is read "is approximately distributed as".

Schruben also gives a variance estimator $\hat{\sigma}^2$ based exclusively on the latter portion of the stochastic process. $\hat{\sigma}^2$ arises from Fishman's autoregressive time series modelling method; Fishman supposes that $\nu\hat{\sigma}^2 \approx \sigma^2\chi^2(\nu)$, where ν must be estimated. Note that $\hat{\sigma}^2$ and $V_{3,1}$ are not necessarily independent.

If we nevertheless assume that $V_{3,1}$ and $\hat{\sigma}^2$ are independent, then

$$F \equiv V_{3,1}/\hat{\sigma}^2 \approx F(3,\nu), \qquad (2\text{-}1)$$

the F distribution with 3 and ν degrees of freedom. Let f be a realization of the random variable F. If we assume that the form of the initialization bias is arbitrary (negative) [positive], then we reject H_0 at the α level if $f < f_{3,\nu,\alpha/2}$ or $> f_{3,\nu,1-\alpha/2}$ ($f > f_{3,\nu,1-\alpha}$) [$f^- > f_{3,\nu,1-\alpha}$], where $f_{3,\nu,\gamma}$ is the upper γ quantile of the F(3,ν) distribution and f^- is the realization of F resulting from the process $-X_1,-X_2,\ldots,-X_n$.

The above testing procedure appears to work well for the battery of simulated systems studied in Schruben (1982). However, the test may not perform adequately if the simulation run is very short (since the test is asymptotic) or if initialization bias pervades the entire process (in which case the experimenter should be able to detect the bias visually). There are also a number of problems inherent with $\hat{\sigma}^2$, including the fact that $\hat{\sigma}^2$ and $V_{3,1}$ are not necessarily independent.

Schruben gives an alternative estimator to $\hat{\sigma}^2$: It is suggested

that the process be divided into $b = 2$ batches, each of size $m = n/2$ so that (assuming the batches are "approximately independent"),

$$\frac{\hat{K}_2(m-\hat{K}_2)}{\hat{K}_1(m-\hat{K}_1)}\,\frac{\hat{S}_1^2}{\hat{S}_2^2} = \frac{\dfrac{m\hat{S}_1^2}{3\hat{K}_1(m-\hat{K}_1)}}{\dfrac{m\hat{S}_2^2}{3\hat{K}_2(m-\hat{K}_2)}} \;:\; F(3,3), \tag{2-2}$$

where the subscripts 1 and 2 refer to the estimators from the first and second batches, respectively.

We will investigate the natural generalization of this test statistic in Section 3.

2.2.3 A Test Based on the Area Estimator

Schruben, Singh, and Tierney (1983) works with a *weighted* form of the *area estimator* for variance to test H_0 vs. H_1: $\mu_i = \mu(1-a_i)$ for all i, for some arbitrary, pre-specified constants a_i, $i=1,\ldots,n$. [This test is unrealistic when $\mu = 0$; in this case, an alternative of the form H_1: $\mu_i = \mu + a_i$ could be used.] The authors standardize the entire output series and find that (under certain strong assumptions) the most powerful test is to reject H_0 when the statistic

$$Z \equiv \sum_{k=1}^{n} c_k k S_k$$

is large, where $c_k \equiv a_k - a_{k+1}$. They give several examples and arguments which show that the experimenter can offer "reasonable" choices for the c_k's.

Since $T_n(t) \xrightarrow{\mathcal{D}} B_t$ as $n \to \infty$,

$$Z = \sum_{k=1}^{n} c_k k S_k = \sqrt{n}\,\sigma \sum_{k=1}^{n} c_k T_n(k/n) \,:\, \sqrt{n}\,\sigma \sum_{k=1}^{n} c_k B_{k/n}. \tag{2-3}$$

This implies that:

$$Z \sim \mathrm{Nor}(0, n\sigma^2 v), \text{ where}$$

$$v \equiv \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j [\min(i/n, j/n) - ij/n^2]. \qquad (2\text{-}4)$$

Given the c_k's it is simple to explicitly calculate v [See Goldsman (1984) for additional discussion concerning weighted area variance estimators.]

As in Section 2.2.2, Schruben, Singh, and Tierney calculate another variance estimator $\hat{\sigma}^2$ and suppose that $\nu\hat{\sigma}^2 \approx \sigma^2\chi^2(\nu)$. The same comments and caveats as before still apply to ν and $\hat{\sigma}^2$. Further supposing that Z and $\hat{\sigma}^2$ are independent yields:

$$\frac{Z}{(n\hat{\sigma}^2 v)^{1/2}} = \frac{Z/(n\sigma^2 v)^{1/2}}{(\hat{\sigma}^2/\sigma^2)^{1/2}} \approx t(\nu).$$

Depending on whether we wish to test for arbitrary, negative, or positive initialization bias, the appropriate test should be performed as in Section 2.2.2. The Schruben, Singh, and Tierney test procedure appears to work well for the examples given in their paper. We will generalize this procedure in the next section.

3. A NEW CLASS OF TESTS FOR INITIALIZATION BIAS

In the ensuing discussion, the various variance estimators given in the previous section will be used to construct new tests for initialization bias. Divide $X_1, \ldots, X_n$ into b adjacent batches, each of size m. Variance estimators based on the first b' batches will be compared to the corresponding estimators from the remaining $b-b'$ batches. This comparison is to be accomplished via an F test.

Using the notation and results from Section 2, we have (under

H_0: $\mu_i = \mu$ for all i):

$$Q_{0,b'} \doteq \sigma^2\chi^2(b'-1), \quad 1 < b' < b-1 \quad \text{(classical)},$$

$$Q_{1,b'} \doteq \sigma^2\chi^2(b'), \quad 1 \le b' \le b-1 \quad \text{(area)},$$

$$Q_{2,b'} \doteq \sigma^2\chi^2(2b'-1), \quad 1 < b' < b-1 \quad \text{(combined classical-area)},$$

$$Q_{3,b'} \doteq \sigma^2\chi^2(3b'), \quad 1 \le b' \le b-1 \quad \text{(maximum)},$$

$$Q_{4,b'} \doteq \sigma^2\chi^2(4b'-1), \quad 1 < b' < b-1 \quad \text{(combined classical-maximum)}.$$

By similar reasoning,

$$Q^*_{0,b-b'} \equiv m \sum_{i=b'+1}^{b} [\bar{X}_{i,m} - \frac{1}{b-b'} \sum_{j=b'+1}^{b} \bar{X}_{j,m}]^2 \doteq \sigma^2\chi^2(b-b'-1),$$

$$Q^*_{1,b-b'} \equiv Q_{1,b} - Q_{1,b'} \doteq \sigma^2\chi^2(b-b'),$$

$$Q^*_{2,b-b'} \equiv Q^*_{0,b-b'} + Q^*_{1,b-b'} \doteq \sigma^2\chi^2(2b-2b'-1),$$

$$Q^*_{3,b-b'} \equiv Q_{3,b} - Q_{3,b'} \doteq \sigma^2\chi^2(3b-3b'),$$

$$Q^*_{4,b-b'} \equiv Q^*_{0,b-b'} + Q^*_{3,b-b'} \doteq \sigma^2\chi^2(4b-4b'-1).$$

To condense notation a bit, define:

$$d_{k,p} \equiv \begin{cases} p-1 & \text{for } k=0 \\ p & \text{for } k=1 \\ 2p-1 & \text{for } k=2 \\ 3p & \text{for } k=3 \\ 4p-1 & \text{for } k=4 \end{cases}$$

Then for all k,

$$Q_{k,b'} \doteq \sigma^2\chi^2(d_{k,b'}) \quad \text{and} \quad Q^*_{k,b-b'} \doteq \sigma^2\chi^2(d_{k,b-b'}).$$

Further, define for all k and b:

$$V_{k,b'} \equiv \frac{Q_{k,b'}}{d_{k,b'}} \quad \text{and} \quad V^*_{k,b-b'} \equiv \frac{Q^*_{k,b-b'}}{d_{k,b-b'}} .$$

Clearly, the V's are variance estimators calculated from the first b' batches, and the V^*'s are the analogous estimators from the remaining $b-b'$ batches. Under the assumption of independent batches, $Q_{k_1,b'}$ is independent of $Q^*_{k_2,b-b'}$, for all k_1,k_2.

We could thus consider test statistics of the form (under H_0):

$$\frac{V_{k_1,b'}}{V^*_{k_2,b-b'}} \doteq F(d_{k_1,b'}, d_{k_2,b-b'}). \tag{3-1}$$

One could attempt to use a test statistic of the form $V_{k_1,b'}/\hat{\sigma}^2$; but since the numerator is not necessarily independent of the denominator, this statistic might not be distributed as $F(d_{k_1,b'},\nu)$.

The test statistics from Section 2.2.2 are easily seen to be special cases of the above [N.B. (2-1) and (2-2) have m = n and m = n/2, respectively.]

For simplicity, we will only work with test statistics of the form (under H_0):

$$\frac{V_{k,b'}}{V^*_{k,b-b'}} \doteq F(d_{k,b'}, d_{k,b-b'}).$$

The goal now is to find that combination of k, b', and b which, in some sense, yields the "optimal" test statistic.

4. POWER CALCULATIONS FOR THE NEW TESTS

A reasonable criterion for comparison among tests (with fixed level of significance) is power. Consider

H_0: $\mu_{i,j} \equiv E[X_{i,j}] = \mu$, where $X_{i,j}$ is the j-th observation from batch i; $i=1,\ldots,b$ and $j=1,\ldots,m$.

vs.

H_1: $\mu_{i,j} = \mu(1-a_{i,j})$, where the $a_{i,j}$'s are pre-specified.

We assume that the form of the initialization bias under H_1 is *negative*. (The cases of arbitrary and positive bias are similar.) Then (3-1) implies that we must reject H_0 at level α if

$$V_{k,b'}/V^*_{k,b-b'} > f_{d_{k,b'},d_{k,b-b'},1-\alpha}.$$

We give analytic results for the cases k = 0, 1, and 2 (classical, area, and classical-area). Limited Monte Carlo results for k = 3 and 4 will be given in Section 4.6.

4.1 Classical Batched Means Tests

Suppose that $E[X_{i,j}] = \mu(1-a_{i,j})$ for all i,j. Define:

$$Y_{i,j} \equiv X_{i,j} + \mu a_{i,j} \quad (\text{so } E[Y_{i,j}] = \mu),$$

$$\bar{Y}_{i,j} \equiv \frac{1}{j}\sum_{p=1}^{j} Y_{i,p},$$

$$\bar{a}_{i,j} \equiv \frac{1}{j}\sum_{p=1}^{j} a_{i,p},$$

$$\bar{\bar{a}}_b \equiv \frac{1}{b}\sum_{i=1}^{b} \bar{a}_{i,m},$$

$$\sigma^2 \equiv \lim_{m\to\infty} m\mathrm{Var}(\bar{Y}_{i,m}).$$

Assuming that the $Y_{i,j}$'s satisfy the mild requirements alluded to in Section 2.1, a direct consequence of Theorem 21.1 of Billingsley

(1968) is that $\bar{Y}_{i,m} \sim \mathrm{Nor}(\mu, \sigma^2/m)$. Hence, $\bar{X}_{i,m} \equiv \frac{1}{m}\sum_{j=1}^{m} X_{i,j} \sim \mathrm{Nor}(\mu(1-\bar{a}_{i,m}), \sigma^2/m)$. This leads to the following:

Theorem:

$$m\sum_{i=1}^{b}(\bar{X}_{i,m}-\bar{\bar{X}}_n)^2 \sim \sigma^2\chi^2\left[b-1,\ (\mu^2 m/\sigma^2)\sum_{i=1}^{b}(\bar{a}_{i,m}-\bar{\bar{a}}_b)^2\right],$$

where $\chi^2(d,\delta)$ is the noncentral χ^2 distribution with d degrees of freedom and noncentrality parameter δ.

Proof: See the Appendix. //

The theorem immediately implies that:

$$Q_{0,b'} \sim \sigma^2\chi^2(b'-1,\ \delta_{0,b'}) \quad \text{and} \quad Q^*_{0,b-b'} \sim \sigma^2\chi^2(b-b'-1,\ \epsilon_{0,b-b'}),$$

where $\delta_{0,b'} \equiv (\mu^2 m/\sigma^2)\sum_{i=1}^{b'}\left[\bar{a}_{i,m} - \frac{1}{b'}\sum_{j=1}^{b'}\bar{a}_{j,m}\right]^2$ and

$$\epsilon_{0,b-b'} \equiv (\mu^2 m/\sigma^2)\sum_{i=b'+1}^{b}\left[\bar{a}_{i,m} - \frac{1}{b-b'}\sum_{j=b'+1}^{b}\bar{a}_{j,m}\right]^2.$$

Hence, $R_{0,b'} \equiv V_{0,b'}/V^*_{0,b-b'} \sim F(b'-1, b-b'-1, \delta_{0,b'}, \epsilon_{0,b-b'})$, where $F(d_1,d_2,\gamma_1,\gamma_2)$ is the doubly noncentral F distribution with d_1 and d_2 degrees of freedom and respective noncentrality parameters γ_1 and γ_2. The power of the test is therefore:

$$\Pr\{\text{Reject } H_0 | H_1 \text{ true}\} = \Pr\{R_{0,b'} > f_{b'-1,b-b'-1,1-\alpha}\}.$$

4.2 Area Tests

The standardized time series from the i-th batch, denoted by $\{T_{i,m}(t)\}$, is given by:

$$T_{i,m}(t) \equiv \frac{\lfloor mt \rfloor S_{i,\lfloor mt \rfloor}}{\sigma\sqrt{m}}, \quad t \in [0,1].$$

Thus,

$$T_{i,m}(j/m) = \frac{jS_{i,j}}{\sigma\sqrt{m}} = \frac{j(X_{i,m}-X_{i,j})}{\sigma\sqrt{m}} \sim T'_{i,m}(j/m) + \frac{\mu c_{i,j}}{\sigma\sqrt{m}},$$

where $T'_{i,m}(j/m) \equiv j(\bar{Y}_{i,m}-\bar{Y}_{i,j})/\sigma\sqrt{m}$ and $c_{i,j} \equiv -j(\bar{a}_{i,m}-\bar{a}_{i,j})$, with the $\bar{Y}_{i,j}$'s and $\bar{a}_{i,j}$'s defined as in Section 4.1. Hence,

$$\hat{A}_i = \sum_{j=1}^{m} jS_{i,j} = \sqrt{m}\,\sigma \sum_{j=1}^{m} T'_{i,m}(j/m) + \mu \sum_{j=1}^{m} c_{i,j}.$$

Note that $T'_{i,m}(.)$ converges to a Brownian bridge asymptotically in m. It is now easy to see [cf. Schruben (1983)] that for large m:

$$\mu c_i^* \equiv E[A_i^*] \doteq \mu \sum_{j=1}^{m} c_{i,j} \quad \text{and}$$

$$\sigma^2 v_i^* \equiv \mathrm{Var}(\hat{A}_i) = \mathrm{Var}\left[\sqrt{m}\,\sigma \sum_{j=1}^{m} T'_{i,m}(j/m)\right] \doteq \frac{m^3-m}{12}\,\sigma^2.$$

So as m becomes large, these results imply that $\hat{A}_i \doteq \mathrm{Nor}(\mu c_i^*, \sigma^2 v_i^*)$ for all i. Since $v^* \equiv v_1^* = \ldots = v_b^*$, we have:

$$\frac{1}{v^*}\sum_{i=1}^{b} \hat{A}_i^2 \doteq \sigma^2\chi^2\left[b,\ (\mu^2/\sigma^2 v^*)\sum_{i=1}^{b}(c_i^*)^2\right].$$

Then

$$Q_{1,b'} \doteq \sigma^2\chi^2\left[b',\ (\mu^2/\sigma^2 v^*)\sum_{i=1}^{b'}(c_i^*)^2\right] \quad \text{and}$$

$$Q^*_{1,b-b'} \doteq \sigma^2\chi^2\left[b-b',\ (\mu^2/\sigma^2 v^*)\sum_{i=b'+1}^{b}(c_i^*)^2\right].$$

So

$$R_{1,b'} \equiv \frac{V_{1,b'}}{V^*_{1,b-b'}} \doteq F(b',\ b-b',\ \delta_{1,b'},\ \epsilon_{1,b-b'}),$$

where $\delta_{1,b'}$ and $\epsilon_{1,b-b'}$ are the obvious noncentrality parameters. The power is $\Pr\{R_{1,b'} > f_{b',b-b',1-\alpha}\}$.

4.3 Classical-Area Tests

Clearly,

$$Q_{2,b'} = Q_{0,b'} + Q_{1,b'} \doteq \sigma^2\chi^2(2b'-1,\ \delta_{2,b'}),$$

where

$$\delta_{2,b'} \equiv \delta_{0,b'} + \delta_{1,b'}.$$

Similarly,

$$Q^*_{2,b-b'} = Q^*_{0,b-b'} + Q^*_{1,b-b'} \doteq \sigma^2\chi^2(2(b-b')-1,\ \epsilon_{2,b-b'}),$$

where

$$\epsilon_{2,b-b'} \equiv \epsilon_{0,b-b'} + \epsilon_{1,b-b'}.$$

So

$$R_{2,b'} \equiv \frac{V_{2,b'}}{V^*_{2,b-b'}} \doteq F(2b'-1,\ 2(b-b')-1,\ \delta_{2,b'},\ \epsilon_{2,b-b'}),$$

which has power $\Pr(R_{2,b'} > f_{2b'-1,2(b-b')-1,1-\alpha})$.

4.4 Weighted Area Tests

As an aside, we describe some additional tests for initialization bias which are similar to those in Schruben, Singh, and Tierney (1983). Once again, test H_0: $\mu_i = \mu$ for all i vs. H_1: $\mu_i = \mu(1-a_i)$ for all i. We construct a test statistic based on (2-3) from the first b′ batches of the $X_1,\ldots,X_n$ process. This statistic is compared to $V^*_{k,b-b'}$ in order to test for bias.

By (2-3) and (2-4),

$$\tilde{Z}_{b'} \equiv \sum_{i=1}^{n_{b'}} c_i i(\bar{X}_{n_{b'}} - \bar{X}_i) \doteq \mathrm{Nor}(0, n_{b'}\sigma^2\tilde{v}_{b'}) ,$$

where $n_{b'} \equiv b'm$ (b′ batches of m X_i's each), $c_i \equiv a_i - a_{i+1}$, and

$$\tilde{v}_{b'} \equiv \sum_{i=1}^{n_{b'}} \sum_{j=1}^{n_{b'}} c_i c_j [\min(i/n_{b'}, j/n_{b'}) - ij/n^2_{b'}] .$$

Assuming independent batches, $\tilde{Z}_{b'}$ and $V^*_{k,b-b'}$ are independent; then under H_0,

$$\tilde{Z}_{b'}/(n_{b'}\tilde{v}_{b'}V^*_{k,b-b'})^{1/2} = \frac{\tilde{Z}_{b'}/(n_{b'}\sigma^2\tilde{v}_{b'})^{1/2}}{(V^*_{k,b-b'}/\sigma^2)^{1/2}} \doteq t(d_{k,b-b'}).$$

Power analysis similar to that from subsections 4.1 – 4.3 can now be performed.

4.5 Analytical Comparison of Tests

We can compare the power of the classical, area, and classical-area test for initialization bias. For given b, b′, m, and k (= 0, 1, or 2), consider the alternative hypothesis H_1: $E[X_{i,j}] = \mu(1-a_{i,j})$, i = 1,...,b; j = 1,...,m. By previous work, the statistic $R_{k,b'} \approx F(d_{k,b'}, d_{k,b-b'}, \delta_{k,b'}, \epsilon_{k,b-b'})$ and has power $\Pr\{R_{k,b'} > f_{d_{k,b'},d_{k,b-b'},1-\alpha}\}$.

Since tables for (singly and doubly) nonventral F distributions are not readily available, we approximate the distribution of $R_{k,b'}$ by using the familiar result [cf. Johnson and Kotz (1970), pg. 197] that:

$$F(\nu_1, \nu_2, \gamma_1, \gamma_2) \approx cF(\tilde{\nu}_1, \tilde{\nu}_2),$$

where

$$c = \frac{(\nu_1+\gamma_1)\nu_2}{(\nu_2+\gamma_2)\nu_1}, \quad \tilde{\nu}_1 = \frac{(\nu_1+\gamma_1)^2}{\nu_1+2\gamma_1}, \text{ and } \tilde{\nu}_2 = \frac{(\nu_2+\gamma_2)^2}{\nu_2+2\gamma_2}.$$

This gives:

$$R_{k,b'} \approx \frac{(d_{k,b'}+\delta_{k,b'})d_{k,b-b'}}{(d_{k,b-b'}+\epsilon_{k,b-b'})d_{k,b'}} F\left[\frac{(d_{k,b'}+\delta_{k,b'})^2}{(d_{k,b'}+2\delta_{k,b'})}, \frac{(d_{k,b-b'}+\epsilon_{k,b-b'})^2}{(d_{k,b-b'}+2\epsilon_{k,b-b'})}\right].$$

So the power is approximately:

$$\Pr\left\{ F(\tilde{\nu}_1,\tilde{\nu}_2) > \frac{1}{c} f_{d_{k,b'},d_{k,b-b'},1-\alpha} \right\}, \qquad (*)$$

where c, $\tilde{\nu}_1$, and $\tilde{\nu}_2$ are the appropriate quantities.

Since c, $\tilde{\nu}_1$, and $\tilde{\nu}_2$ are functions of $\delta_{k,b'}$ and $\epsilon_{k,b-b'}$, they are functions of μ^2/σ^2, which is unknown. If we are willing to estimate the value of μ^2/σ^2, that combination of b, b′, m, and k can be found which maximizes (*).

A simple example

Suppose that m = 100, b = 20, and H_1: $E[X_{i,j}] = \mu(1-a_{i,j})$, where

$a_{i,j} = \left[1 - \frac{(i-1)m+j}{bm}\right]^2$ for all i,j. Figures 2 through 7 give plots of (*) as a function of μ^2/σ^2 for b' = 2,5,8,10,15, and 18, respectively. Each of the six figures illustrate plots for the k = 0, 1, and 2 cases with the level α = 0.05.

Suppose we fix b' = 2. From Figure 2, it is easily seen that for all values of $\mu^2/\sigma^2 \in (0,5]$, the power for the specific alternative hypothesis is maximized when k = 1 (area test).

Alternatively, fix b' = 8 and consult Figure 4. We see that the power (*) is maximized for $\mu^2/\sigma^2 \in (0,.5)$ by k = 0 (classical test) and for $\mu^2/\sigma^2 \in [.5,5]$ by k = 2 (combined classical-area test). Finally, Figure 7 reveals that when b' = 18, (*) is maximized for $\mu^2/\sigma^2 \in (0,5]$ by k = 2.

The above results are to be expected if we note that the classical and combined classical-maximum variance estimators are more sensitive to between batch changes in mean than is the area variance estimator.

4.6 Empirical Comparison of Tests: An Example

We empirically compare the power of the classical, area, classical-area, maximum, and classical-maximum initialization bias tests for the following example:

Consider the stationary first order autoregressive process $X_1,\ldots,X_n$, with

$$X_i = \beta X_{i-1} + \varepsilon_i, \text{ where } \varepsilon_i \sim \text{iid Nor}(0,1-\beta^2).$$

Let $Y_{i,j} \equiv X_{i,j} + \mu(1-a_{i,j})$, i=1,...,b; j=1,...,m. We test H_0: $E[Y_{i,j}] = \mu$ for all i,j against H_1: $E[Y_{i,j}] = \mu(1-a_{i,j})$. Suppose

we take $\mu = 1$ and m, b, and $a_{i,j}$ as in the previous example. Let the level $\alpha = 0.05$.

For each of various values of b' and AR(1) coefficient β, we ran 500 independent Monte Carlo experiments; we applied the five initialization bias tests ($k = 0,1,2,3,4$) to each experiment. For given b', β, and k, the estimated power = (number of experiments for which H_0 is rejected)/500. Table 1 summarizes these results.

Remarks:

(1) For this example, the maximum test ($k = 3$) or the classical-maximum test ($k = 4$) appear to be the most powerful.

(2) For this example, it turns out [cf. the proof of Result 5-5 in Goldsman (1984)] that $\sigma^2 = (1+\beta)/(1-\beta)$; so $\mu^2/\sigma^2 = (1-\beta)/(1+\beta)$. We then see that the $k = 0$, 1, and 2 entries in Table 1 closely match the results in Figures 2 through 7.

5. CONCLUSIONS

We have constructed a general family of tests for detecting initialization bias in a simulated process: The process is divided into two adjacent, non-overlapping portions. A variance estimate is then calculated from each of the two portions. If these estimates are deemed to be significantly different, then bias is said to be present. Of course, a good deal of the simulation literature deals with the question of variance estimation; our estimators are primarily rooted in the recent standardized time series work, but use of other variance estimators is straightforward.

A criterion of desirability for a particular initialization bias

test is that it be powerful. (We were able to derive analytic power results for the classical, area, and combined classical-area cases.) However, as the examples of the previous section show, there is not necessarily a choice of k, b, and b′ which yields the most powerful test in all situations. Therefore, the authors are currently addressing this problem of determining the best k, b, and b′ for general processes.

Acknowledgement: It is our pleasure to thank Prof. Lionel Weiss for his suggestions concerning the proof of the theorem in Section 4.1.

References

Andrews, R.W. and T.J. Schriber (1982). Two ARMA-based confidence interval procedures for the analysis of simulation output. Working paper #304, Graduate School of Business Administration, The University of Michigan, Ann Arbor, Michigan.

Billingsley, P. (1968). Convergence of Probability Measures. John Wiley and Sons, New York.

Fishman, G.S. (1971). Estimating sample size in computer simulation experiments. Management Science, 18, 21-38.

Fishman, G.S. (1973). Concepts and Methods in Discrete Event Digital Simulation. John Wiley and Sons, New York.

Fishman, G.S. (1978). Principles of Discrete Event Simulation. John Wiley and Sons, New York.

Goldsman, D. (1984). On Using Standardized Time Series to Analyze Stochastic Processes, Ph.D. Thesis, School of ORIE, Cornell Univ., Ithaca, New York.

Heidelberger, P. and P.D. Welch (1981). A spectral method for confidence interval generation and run-length control in simulators. Comm. ACM, 24, 233-245.

Johnson, N.L. and S. Kotz (1970). Distributions in Statistics - Continuous Univariate Distributions - 2. John Wiley and Sons, New York.

Rohatgi, V.K. (1976). An Introduction to Probability Theory and Mathematical Statistics. John Wiley and Sons, New York.

Schruben, L. (1981). Control of initialization bias in multivariate simulation response. Comm. ACM, 24, 246-252.

Schruben, L. (1982). Detecting initialization bias in simulation output Operations Research, 30, 569-590.

Schruben, L. (1983). Confidence interval estimation using standardized time series. Operations Research, 31, 1090-1108.

Schruben, L. and D. Goldsman (1984). Initialization effects in computer simulation experiments. Tech. Report 594, School of ORIE, Cornell Univ., Ithaca, New York.

Schruben, L., H. Singh, and L. Tierney (1983). Optimal tests for initialization bias in simulation output. Operations Research, 31, 1167-1178.

Snell, M.K. and L. Schruben (1979). The use of weighting functions to correct for initialization bias in the AR(1) process. Tech. Report 395, School of ORIE, Cornell University, Ithaca, New York.

Wilson, J. and A.A.B. Pritsker (1978). A survey of research on the simulation startup problem. Simulation, 31, 55-58.

Appendix

We prove the theorem in Section 4.1; viz., if

$$\bar{X}_{i,m} \approx \text{Nor}(\mu(1-\bar{a}_{i,m}), \frac{\sigma^2}{m}),$$

then

$$m \sum_{i=1}^{b} (\bar{X}_{i,m} - \bar{\bar{X}}_n)^2 \approx \sigma^2 \chi^2\left(b-1, \frac{\mu^2 m}{\sigma^2} \sum_{i=1}^{b} (\bar{a}_{i,m} - \bar{\bar{a}}_b)^2\right)$$

Proof: Consider independent $U_i \sim \text{Nor}(\nu_i, \tau^2)$, $i=1,\ldots,b$. There exists a nonrandom orthogonal bxb matrix H whose last row is $(b^{-1/2}, b^{-1/2}, \ldots, b^{-1/2})$. (An example of such a matrix is easy to construct.) Define $\underset{\sim}{Z} \equiv (Z_1,\ldots,Z_b)$ by $\underset{\sim}{Z} \equiv \underset{\sim}{U}H^T$. Then $Z_b = \sqrt{b}\ \bar{U}_b$, where $\bar{U}_b \equiv \frac{1}{b} \sum_{i=1}^{b} U_i$.

Since H is orthogonal, $\underset{\sim}{Z}\underset{\sim}{Z}^T = \underset{\sim}{U}H^T(\underset{\sim}{U}H^T)^T = \underset{\sim}{U}H^T H\underset{\sim}{U}^T = \underset{\sim}{U}\underset{\sim}{U}^T$; so $\sum_{i=1}^{b} Z_i^2 = \sum_{i=1}^{b} U_i^2$. Thus,

$$S^2 \equiv \sum_{i=1}^{b} (U_i - \bar{U}_b)^2 = \sum_{i=1}^{b} U_i^2 - b\bar{U}_b^2 = \sum_{i=1}^{b} Z_i^2 - Z_b^2 = \sum_{i=1}^{b-1} Z_i^2.$$

Since H is a nonrandom matrix, $E[\underset{\sim}{Z}] = E[\underset{\sim}{U}H^T] = \underset{\sim}{\nu}H^T$. The variance-covariance matrix of $\underset{\sim}{Z}$ is given by:

$$D[\underset{\sim}{Z}] = E[(\underset{\sim}{Z}-\underset{\sim}{\nu}H^T)^T(\underset{\sim}{Z}-\underset{\sim}{\nu}H^T)]$$

$$= E[(\underset{\sim}{U}H^T-\underset{\sim}{\nu}H^T)^T(\underset{\sim}{U}H^T-\underset{\sim}{\nu}H^T)]$$

$$= E[H(\underset{\sim}{U}-\underset{\sim}{\nu})^T(\underset{\sim}{U}-\underset{\sim}{\nu})H^T]$$

$$= HE[(\underset{\sim}{U}-\underset{\sim}{\nu})^T(\underset{\sim}{U}-\underset{\sim}{\nu})]H^T = HI\tau^2H^T,$$

because the U_i's are independent $\mathrm{Nor}(\nu_i,\tau^2)$ random variables. So $D[\underset{\sim}{Z}] = I\tau^2$.

Since each Z_i is a linear combination of normal random variables, we have:

$$\underset{\sim}{Z} \sim \mathrm{Nor}_b(\underset{\sim}{\nu}H^T, I\tau^2).$$

I.e., the Z_i's are independent normal random variables each with variance τ^2. Thus,

$$\frac{S^2}{\tau^2} = \sum_{i=1}^{b-1} (Z_i/\tau)^2 \sim \chi^2(b-1,\delta),$$

where δ is computed as follows: It $Q \sim \chi^2(b-1,\delta)$, then $E[Q] = b-1+\delta$ [see Rohatgi (1976), pg. 315]. This yields:

$$E[S^2/\tau^2] = \frac{1}{\tau^2}\{E[\sum_{i=1}^{b} U_i^2] - bE[\bar{U}_b^2]\}$$

$$= \frac{1}{\tau^2}\{\sum_{i=1}^{b}[Var(U_i) + (EU_i)^2] - b[Var(\bar{U}_b) + (E\bar{U}_b)^2]\}$$

$$= \frac{1}{\tau^2}\{b\tau^2 + \sum_{i=1}^{b}\nu_i^2 - \tau^2 - b\bar{\nu}_b^2\}, \text{ where } \bar{\nu}_b \equiv \frac{1}{b}\sum_{i=1}^{b}\nu_i$$

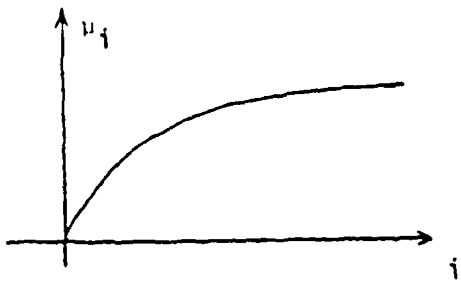
$$= b-1 + \frac{1}{\tau^2}\sum_{i=1}^{b}(\nu_i-\bar{\nu}_b)^2.$$

So

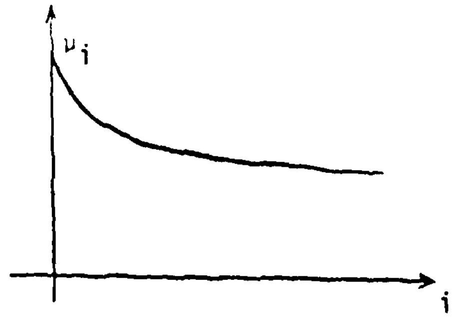
$$\delta = \frac{1}{\tau^2}\sum_{i=1}^{b}(\nu_i-\bar{\nu}_b)^2.$$

The proof is completed if we identify σ^2/m with τ^2, $\bar{X}_{i,m}$ with U_i, and $\mu(1-\bar{a}_{i,m})$ with ν_i, $i = 1,\ldots,b$. //

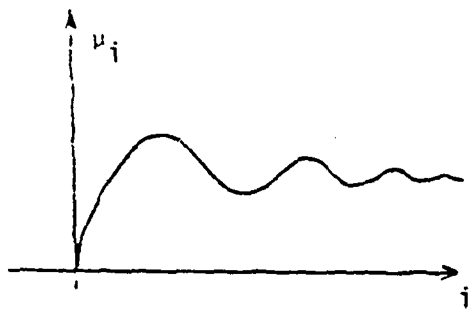
An alternative proof is given in Appendix A.3 of Goldsman (1984).



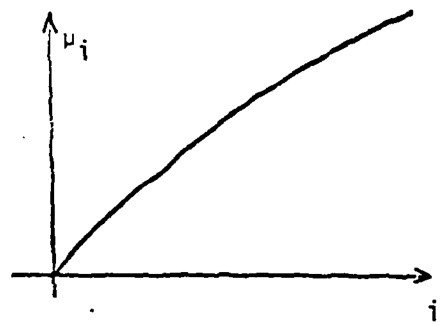
(a) negative bias



(b) positive bias



(c) damping bias



(d) a transient mean process which does not appear to be approaching steady state

Figure 1: Various transient mean processes (drawn as continuous functions for clarity).

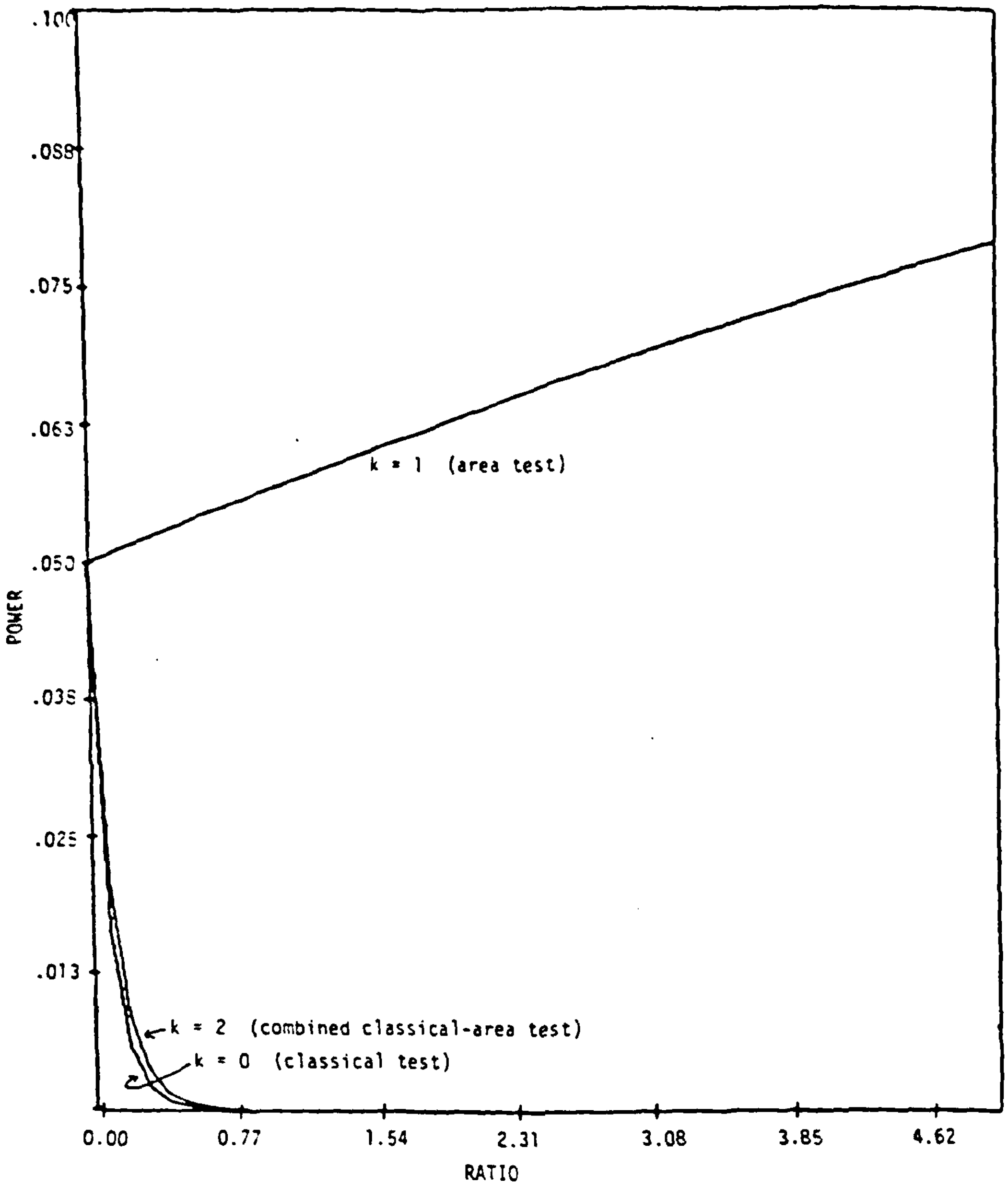


Figure 2: The power (*) as a function of the ratio μ^2/c^2 for the example described in the text for $b' = 2$ and $k = 0,1,2$. N.B. The ordinate axis runs from 0.00 to 0.10.

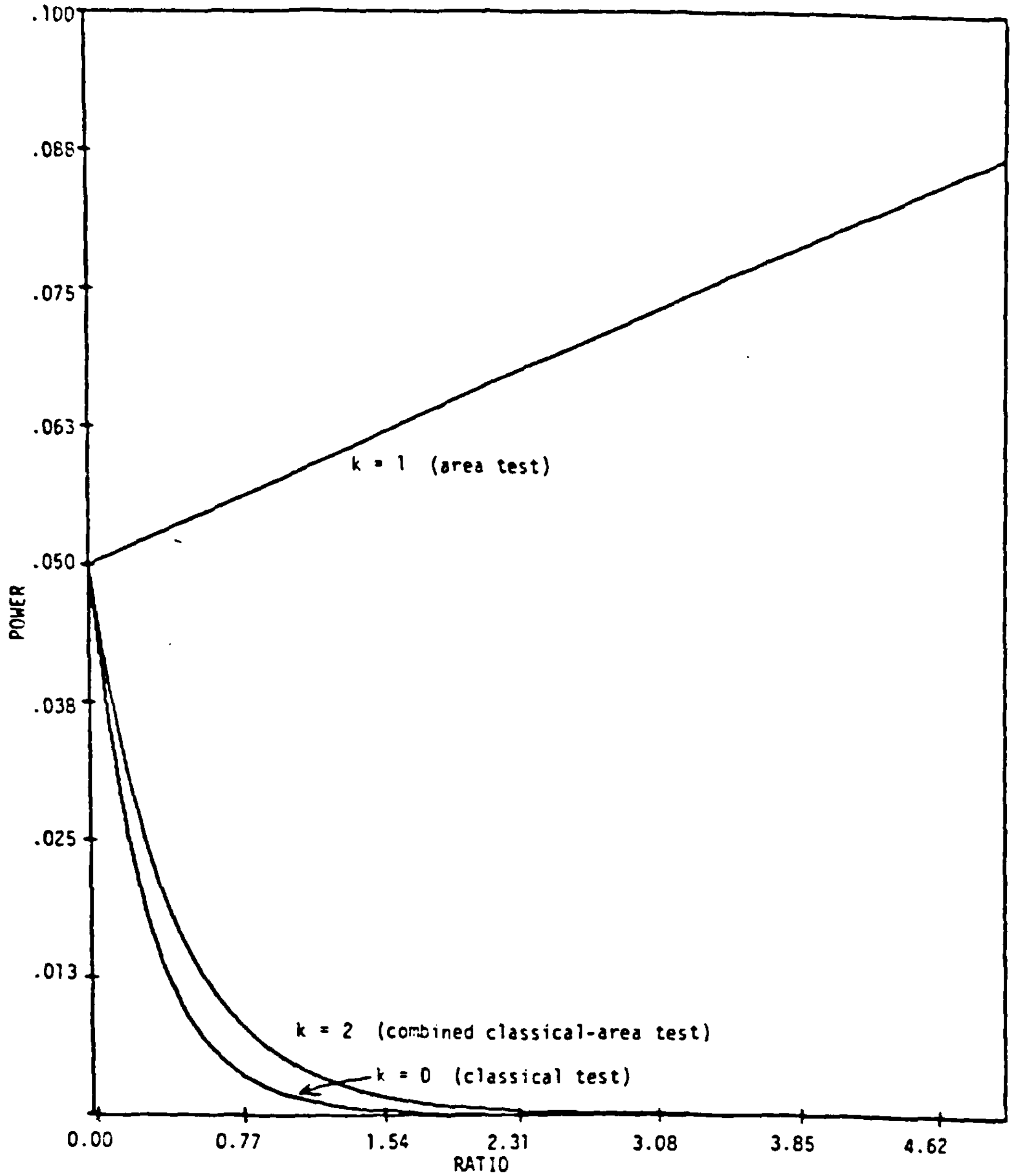
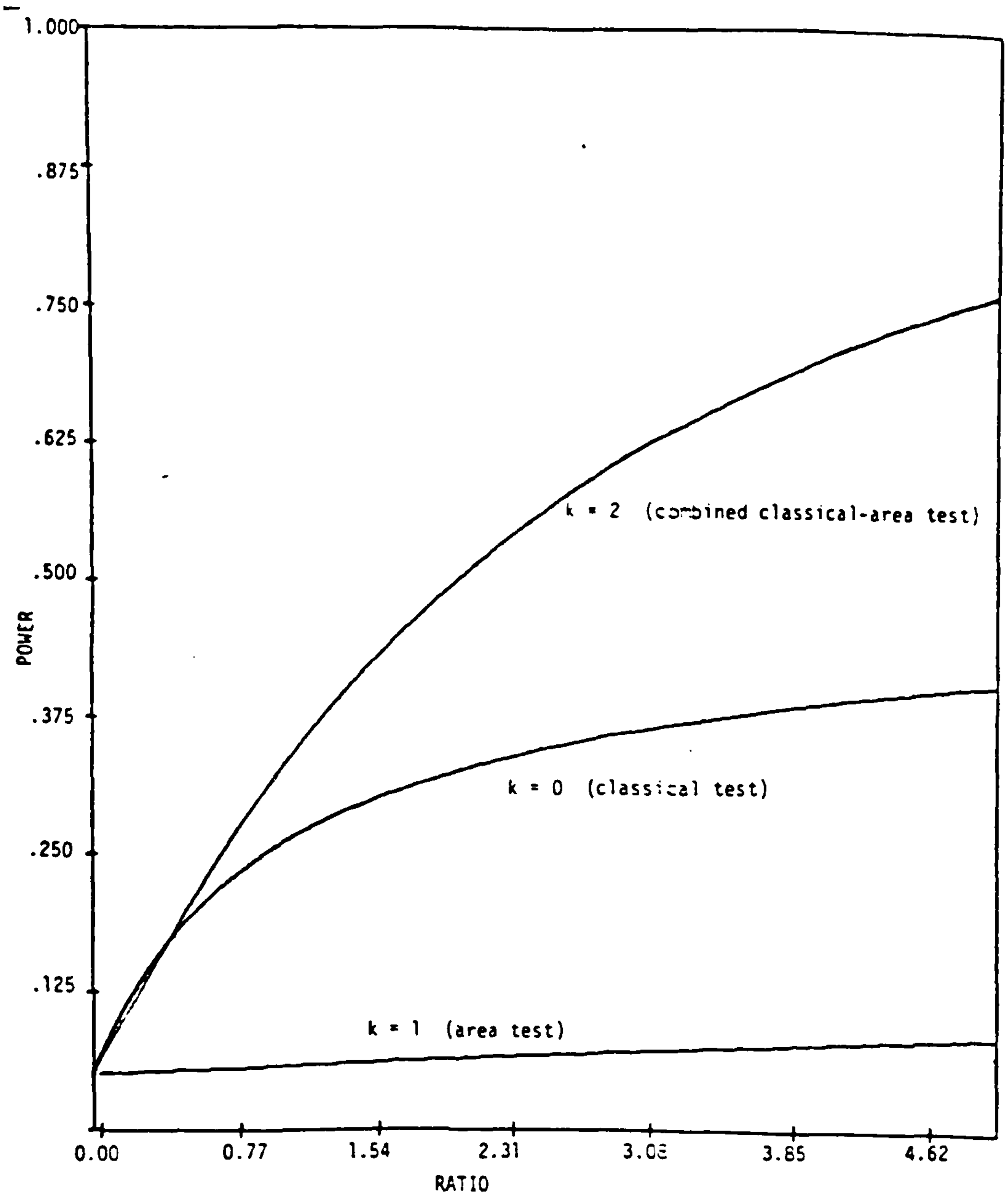


Figure 3: The power (*) as a function of the ratio μ^2/σ^2 for the example described in the text for b' = 5 and k = 0,1,2. N.B. The ordinate axis runs from 0.00 to 0.10.



<u>Figure 4</u>: The power (*) as a function of the ratio [illegible]$^2/\sigma^2$ for the example described in the text for $b' = 8$ and $k = 0,1,2$.

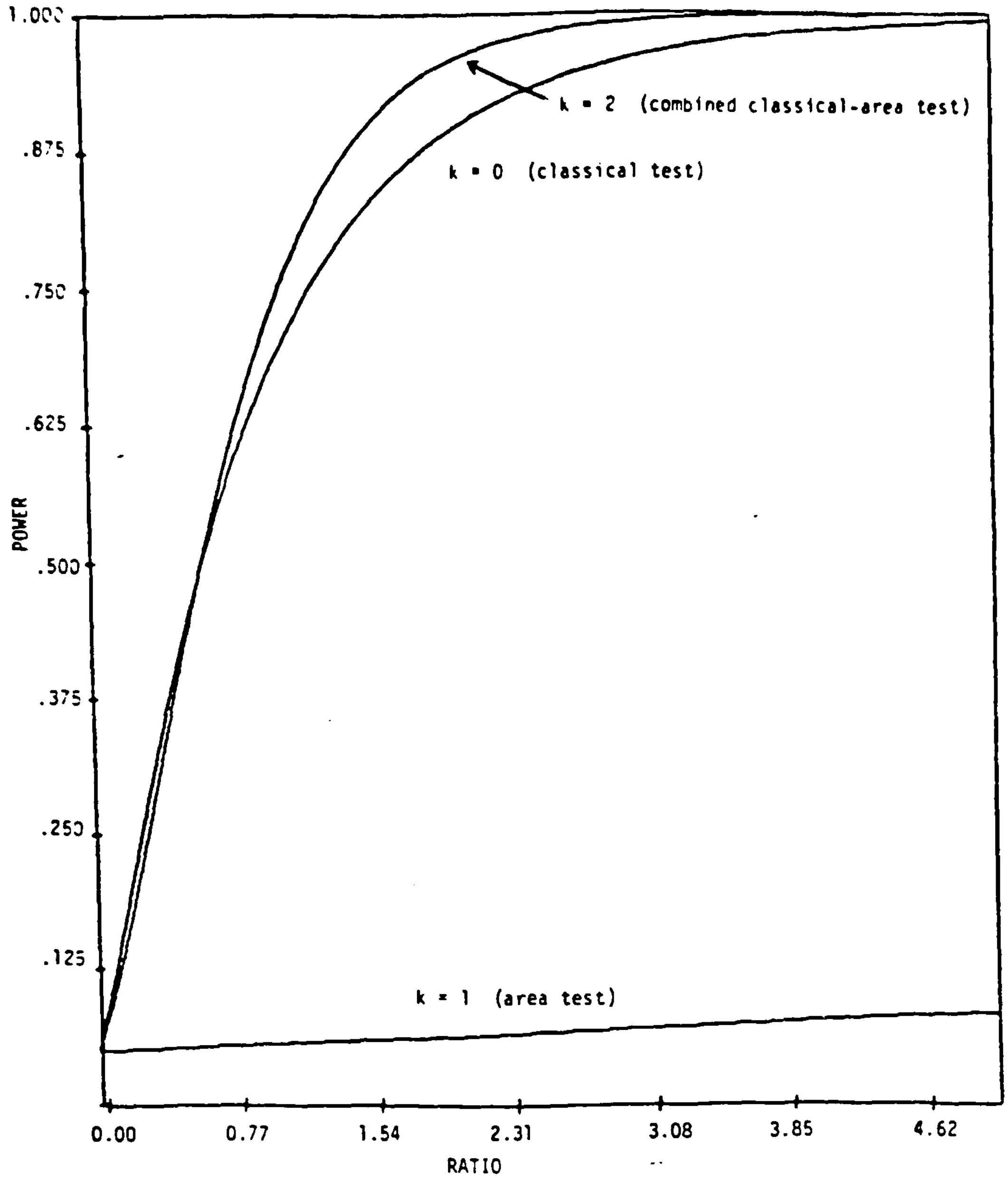


Figure 5: The power (*) as a function of the ratio μ^2/σ^2 for the example described in the text for b' = 10 and k = 0,1,2.

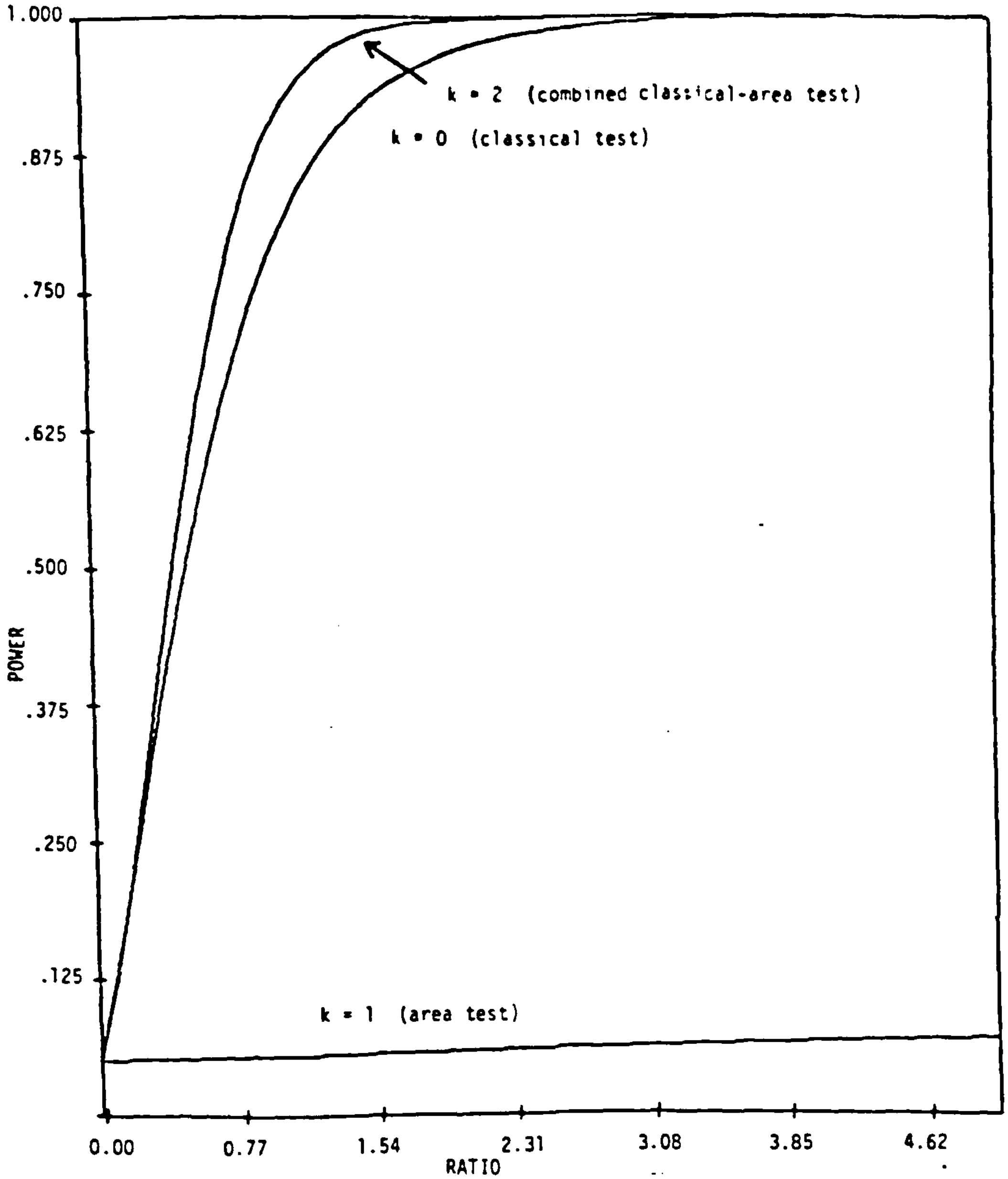


Figure 6: The power (*) as a function of the ratio [illegible]^2/[illegible]^2 for the example described in the text for b' = [illegible] and k = 0,1,2.

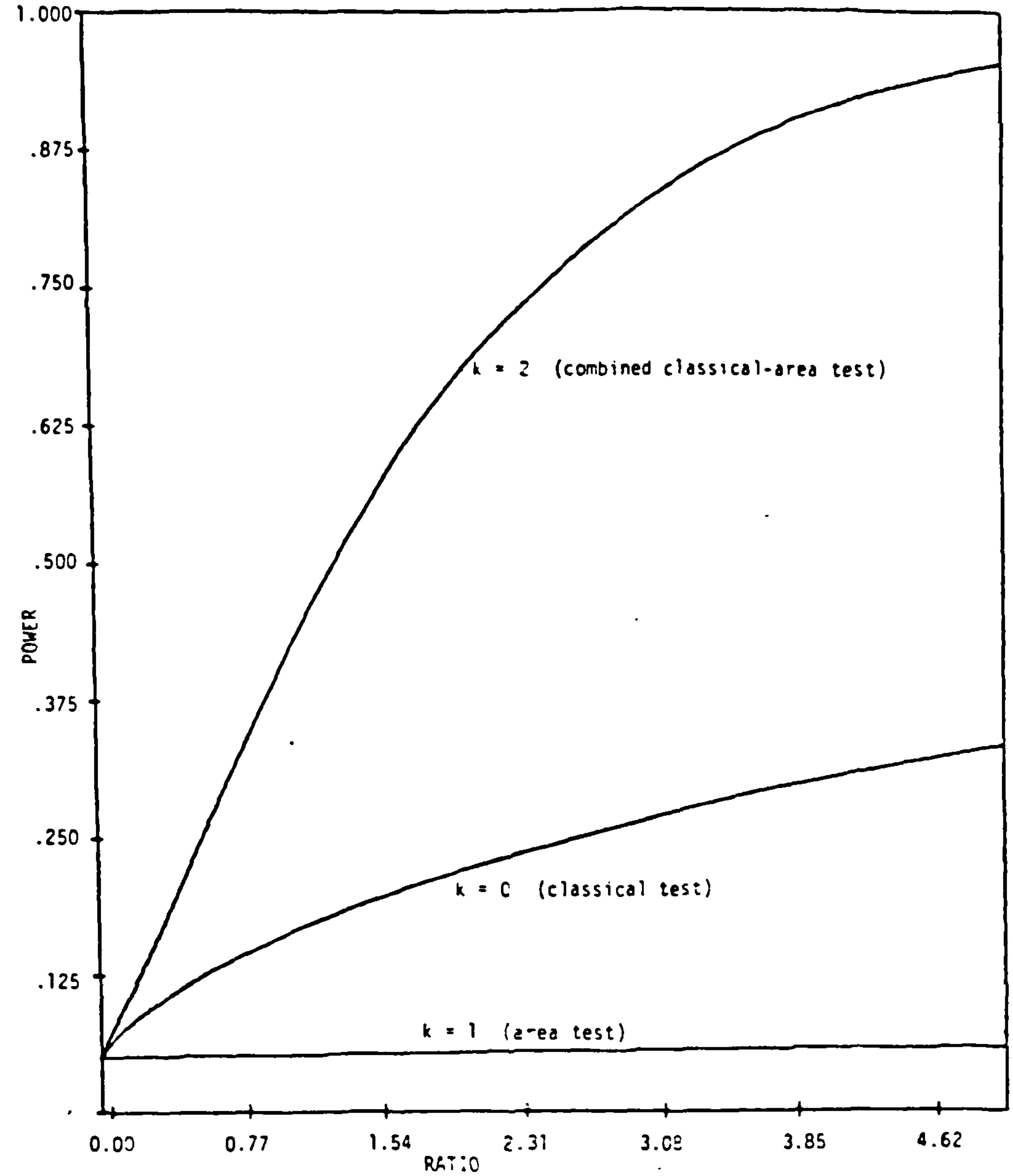


Figure 7: The power (*) as a function of the ratio $\_^2/c^2$ for the example described in the text for b' = 18 and k = 0,1,2.

	b' = 2	5	8	10	15	18
α = -0.5						
k = 0	.000	.000	.396	.970	.992	.270
k = 1	.058	.060	.092	.088	.066	.066
k = 2	.000	.000	.626	.996	1.000	.812
k = 3	.056	.082	.116	.086	.108	.096
k = 4	.000	.000	.710	.996	1.000	.948
α = 0.0		#				
k = 0	.000	.002	.260	.658	.762	.170
k = 1	.070	.076	.050	.064	.044	.064
k = 2	.000	.005	.304	.732	.886	.422
k = 3	.090	.108	.076	.098	.110	.080
k = 4	.000	.012	.368	.742	.942	.652
α = 0.5						
k = 0	.006	.016	.164	.312	.370	.092
k = 1	.050	.054	.038	.044	.060	.068
k = 2	.006	.022	.166	.290	.436	.156
k = 3	.066	.108	.132	.122	.124	.110
k = 4	.010	.052	.208	.364	.532	.322

Table 1

Estimated power of tests for initialization bias for a shifted AR(1) process. For given b' and α, the five table entries are based on the same 500 (# = 1000) independent experiments. The level of significance = 0.05. See the text for details.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS Unrestricted
2a. SECURITY CLASSIFICATION AUTHORITY Office of Naval Research	3. DISTRIBUTION/AVAILABILITY OF REPORT Unlimited Distribution
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE Not applicable	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report No. J-84-16	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Cornell University / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research
6c. ADDRESS (City, State and ZIP Code) Ithaca, NY 14853	7b. ADDRESS (City, State and ZIP Code) Arlington, VA 22217
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research / 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-81-K-0037
8c. ADDRESS (City, State and ZIP Code) Arlington, VA 22217	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO.
11. TITLE (Include Security Classification) Tests for initialization bias in computer simulation experiments (unclassified)	

12. PERSONAL AUTHOR(S)
David Goldsman and Lee Schruben

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Interim	FROM Mar 31 84 TO Aug 25 85	1984 December	31

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Although many of the rules for detecting and dealing with *initialization bias* in computer simulation experiments are easy to understand and implement, they are nonetheless heuristic. The current paper uses the theory of standardized time series to construct tests which (under certain conditions) detect "significant" initialization bias in a process. Previous tests for initialization bias can be viewed as *special cases* of the general family of tests to be presented here.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Lee W. Schruben	22b. TELEPHONE NUMBER (Include Area Code) (607) 256-4856 / 22c. OFFICE SYMBOL

END

FILMED

1-86

DTIC